



**GRT INSTITUTE OF
ENGINEERING AND
TECHNOLOGY, Tiruttani**



**1103-GRT INSTITUTE OF ENGINEERING AND
TECHNOLOGY**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROJECT TITLE

Future sales prediction

Mohanapriya.S

3rd yr, 5th sem

Reg no:110321104029

mohanapriyas144@gmail.com

PHASE3: FUTURE SALES PREDICTION

Future sales prediction, also known as sales forecasting, is the process of estimating a company's future sales based on historical data, market trends, and various analytical techniques. Accurate sales predictions are crucial for businesses as they help in making informed decisions regarding inventory management, resource allocation, budget planning, and overall strategy. Predicting future sales is a common and valuable task in various industries, including retail, e-commerce, and more. To implement a sales prediction project, you can use various techniques and tools.

HERE'S A STEP-BY-STEP GUIDE ON HOW TO IMPLEMENT CUSTOMER SEGMENTATION USING DATA SCIENCE:

1. DATA COLLECTION:

- Gather relevant data on your customers. This data can include demographic information (age, gender, location), behavioral data (purchase history, browsing behavior, engagement with your products or services), and psychographic data (lifestyle, interests, values). You may collect this data through your CRM system, website analytics, surveys, or other sources.

2. DATA PREPROCESSING:

- Clean the data to ensure it's accurate and consistent. This involves handling missing values, removing duplicates, and dealing with outliers. You may also need to transform data into a suitable format for analysis.

3. CHOOSE A PREDICTION TECHNIQUE:

- There are several methods to segment customers, and the choice depends on your data and business goals:

Time Series Forecasting:

Time series forecasting methods, such as ARIMA (AutoRegressive Integrated Moving Average), Exponential Smoothing, and Prophet, are specifically designed for predicting future values based on historical time-ordered data.

Linear Regression:

Linear regression can be used to model the relationship between various factors (features) and sales. You can include relevant predictors such as historical sales, advertising spend, seasonality, and economic indicators to make sales predictions.

Machine Learning Models:

Supervised machine learning models, including decision trees, random forests, support vector machines, and gradient boosting, can be used to capture complex relationships in the data.

Facebook Prophet:

Prophet is a forecasting tool developed by Facebook that is specifically designed for time series data. It can handle seasonality, holidays, and missing data.

The choice of technique depends on the specific characteristics of your sales data, such as seasonality, trends, external factors, and the size of your dataset.

4. DATA MODELING:

- Apply the chosen prediction technique to your data. This will result in the prediction of the future sales.

5. IMPLEMENTATION:

- Load and preprocess your sales data.
- Split the data into features and the target variable.
- Split the data into training and testing sets for model evaluation.
- Create a linear regression model and train it on the training data.
- Use the model to make predictions on the test set and evaluate its performance.
- Visualize the results if needed.
- Finally, you can use the trained model to make predictions for future sales based on your future feature values.

In addition to these steps, it's essential to keep the business goals in mind when implementing the future sales prediction. Each business is unique, and the prediction strategy should align with the specific objectives.

3.1 DATASET AND ITS DETAIL EXPLANATION AND IMPLEMENTATION

This dataset is designed for forecasting future sales in a retail context. It contains historical sales data along with various features that can be used to build predictive models.

When working with a dataset, it's essential to perform data exploration and preprocessing to better understand its characteristics and prepare it for modeling.

We've got the data set in the website called Kaggle(www.kaggle.com/data)

The data set which is respective to our project is sales.csv

(<https://www.kaggle.com/datasets/chakradharmattapalli/future-sales-prediction>)

The data set having the 4 columns named TV, Radio, Newspaper, Sales and having 200 rows of data. (numerical values).

The dataset given here contains the data about the sales of the product. The dataset is about the advertising cost incurred by the business on various advertising platforms. Below is the description of all the columns in the dataset:

TV: Advertising cost spent in dollars for advertising on TV;

Radio: Advertising cost spent in dollars for advertising on Radio;

Newspaper: Advertising cost spent in dollars for advertising on Newspaper;

Sales: Number of units sold

So, in the above dataset, the sales of the product depend on the advertisement cost of the product.

To predict future sales, you can create a dataset with columns for TV advertising spending, radio advertising spending, newspaper advertising spending, and sales.

3.2 BEGIN THE PROJECT BY LOADING THE DATASET

THE SAMPLE DATA SET:

TV	Radio	Newspaper	Sales
203.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12
151.5	41.3	58.5	16.5
180.8	10.8	58.4	17.9

1.IMPORT NECESSARY LIBRARIES

To work with the data and build a predictive model for future sales based on the TV, radio, and newspaper advertising expenditures, you will typically use Python and several libraries.

Here are the key libraries you may need and how to download them: make sure to activate your virtual environment first and then run the pip install commands within that environment.

- Numpy
- Pandas
- Matplotlib and seaborn
- Sci-kit Learn(sk learn)

Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Load the data

```
df = pd.read_csv('C:\Users\priya\Downloads\archive.zip')
```

Explore the data

```
# Display the first few rows of the dataset
print(df.head())

# Check the basic statistics of the dataset
print(df.describe())

# Check for missing values
```

```
print(df.isnull().sum())
```

```
# Check the data types of each column
```

```
print(df.dtypes)
```

data visualization

```
# Create a scatter plot
plt.figure(figsize=(8, 6))
plt.scatter(data['TV'], data['Sales'], c='b', marker='o', label='TV vs. Sales')
plt.title('Scatter Plot of TV Advertising vs. Sales')
plt.xlabel('TV Advertising Budget')
plt.ylabel('Sales')
plt.legend()
plt.grid(True)
plt.show()
```

Model Building

```
X = df[['TV', 'Radio', 'Newspaper']]
y = df['Sales']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
#for future sales prediction
model = LinearRegression()
model.fit(X_train, y_train)
model.fit(X_train, y_train)
#features = [[TV, Radio, Newspaper]]
features = np.array([[230.1, 37.8, 69.2]])
print(model.predict(features))
```

Model Evaluation

```
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

3.3 PREPROCESSING THE DATASET

Import library

```
import pandas as pd
# for manipulation
(the given data)
```

Code

```
x = np.array(data.drop(["Sales"], 1))  
y = np.array(data["Sales"])  
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2, random_state=42)
```

Visualization

```
#relationship between the amount spent on advertising on TV and units sold  
import plotly.express as px  
import plotly.graph_objects as go  
figure = px.scatter(data_frame = data, x="Sales", y="TV", size="TV", trendline="ols")  
figure.show()
```

3.4 PERFORMING DIFFERENT ANALYSIS NEEDED

Predicting future sales is essential for businesses to make informed decisions and plan their strategies effectively. Several analyses and techniques can be performed to improve the accuracy of sales predictions. Here are different types of analysis needed for future sales prediction:

Time Series Analysis:

Historical Sales Data Analysis: Analyze past sales data to identify trends, seasonality, and any patterns over time.

Decomposition: Decompose time series data into trend, seasonality, and residual components to understand underlying patterns.

Market and Industry Analysis:

Market Research: Examine the market conditions, including competitors, economic factors, and customer preferences.

Industry Trends: Identify trends and developments in the industry that may affect sales.

Customer Segmentation:

Customer Demographics: Analyze customer characteristics such as age, gender, location, and buying behavior to identify different customer segments.

Customer Lifetime Value:

Calculate the lifetime value of different customer segments to prioritize marketing efforts.

Product Analysis:

Product Performance: Evaluate the sales performance of individual products or product categories.

Product Life Cycle: Determine where products are in their life cycle to anticipate future sales.

Seasonal Analysis:

Seasonal Effects: Study the impact of seasons, holidays, and special events on sales patterns.

Seasonal Promotions: Analyze the effectiveness of seasonal promotions on sales.

Marketing and Advertising Analysis:

Marketing Campaign Performance: Evaluate the impact of marketing campaigns on sales.

Advertising ROI: Measure the return on investment of advertising efforts.

Economic Indicators:

Economic Factors: Consider economic indicators like GDP, inflation, and unemployment rates that can influence consumer spending.

Customer Feedback and Surveys:

Customer Feedback Analysis: Analyze customer feedback and surveys to understand their preferences and concerns.

Net Promoter Score (NPS): Monitor NPS scores to gauge customer satisfaction and loyalty.

Machine Learning and Predictive Modeling:

Regression Analysis: Use regression models to analyze the relationship between various factors and sales.

Machine Learning Algorithms: Employ machine learning techniques such as decision trees, random forests, or neural networks for predictive modeling.

Inventory and Supply Chain Analysis:

Inventory Levels: Analyze inventory levels and stockouts to ensure products are available when customers want to purchase.

Supply Chain Efficiency: Optimize the supply chain to meet demand efficiently.

Geographic Analysis:

Geographic Sales Patterns: Analyze sales data by region to identify geographic variations in sales.

Expansion Opportunities: Identify potential regions for expansion based on sales data.

Customer Churn Analysis:

Churn Rate: Analyze customer churn rate to understand customer retention and attrition.

External Data Sources:

External Data Integration: Incorporate external data sources like weather data, social media trends, or demographic information to enhance predictions.

Scenario Analysis:

What-If Scenarios: Analyze different scenarios based on changing market conditions, pricing strategies, or marketing investments.

Forecast Accuracy Assessment:

Evaluate the accuracy of past forecasts to continually refine and improve prediction models.

Remember that the specific analyses needed for sales prediction will vary depending on the industry, business size, and the availability of data. It's essential to leverage a combination of these methods and adapt them to your specific business context to make more accurate sales predictions.